JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **158**, 396–406 (1991)

Average Cost Markov Decision Processes: Optimality Conditions*

O. HERNÁNDEZ-LERMA

Depto. de Matemáticas, CINVESTAV-IPN, Apdo. Postal 14-740, México, D.F. 07000, México

AND

J. C. HENNET AND J. B. LASSERRE

LAAS-CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse Cédex, France

Submitted by E. Stanley Lee

Received October 24, 1989

1. INTRODUCTION

We are concerned in this paper with discrete-time Markov Decision Processes (MDPs) with Borel state and action spaces X and A, respectively, and the long run expected *average cost* criterion. When X is a *denumerable* set, many necessary and/or sufficient conditions for the existence of optimal control policies are known. However, when X is a *Borel space* (i.e., a Borel subset of a complete separable metric space), most of the available results impose on the MDP very restrictive topological conditions (e.g., compactness) and/or strong recurrence assumptions (such as Doeblin's condition); see, e.g., [4, 9, 12] and their references. Another related work is [7] where we have studied MDPs from the viewpoint of the recurrence (or ergodicity) properties of the state process. In the present paper, however, we are concerned with the existence of *average optimal* policies by looking at (static) optimization problems (see condition C5 in Section 4) related—in some cases equivalent—to the existence of a bounded solution to the so-called Optimality Equation (see C4 in Section 4). These optimization problems are "dual" in the sense that, under appropriate conditions, the existence of an optimal solution to one of the problems implies existence of an optimal

* This work was partially supported by a joint CONACYT (México)–CNRS (France) research program. The research of the first author was also supported in part by the TWAS under Grant RG-MP 898-152.

0022-247X/91 $3.00

solution to the other(s) and, moreover, the corresponding optimal values of the problems are equal. More generally, feasible solutions to one of the problems provide bounds for the other. This approach is more or less standard when X and A are both *finite* sets, as in [1] and references therein, but in a more general setting it has been followed only by Yamada [12], who assumes that X is a *compact* subset of R^n and that the transition law has a density which satisfies a certain "positivity" condition (see (A1) in Remark 3.2 below). Here, we obtain results similar to those in [1, 12] in the setting of general Borel spaces, and furthermore, our "static" problems have formally a simpler form. Also, using the concept of "opportunity cost" introduced by Flynn [2, 3], we show that a stationary policy determined from the optimality equation is *strong* average optimal (see Definition 2.2).

Our main results are presented in Section 4; they roughly consist of relations between several ergodicity and optimality conditions introduced in Section 3. We begin in Section 2 by presenting the Markov decision model and the optimality criteria we are interested in.

2. Preliminaries

We will use the following notation. A *Borel space* X (i.e., a Borel subset of a complete separable metric space) is always endowed with the Borel sigma-algebra $\mathscr{B}(X)$. $P(X)$ and $B(X)$ denote the space of probability measures on X and the space of real-valued measurable bounded functions on X, respectively. If $v \in B(X)$, $\|v\|$ denotes its supremum norm, whereas if μ is a finite signed measure on X, $\|\mu\|$ stands for the total variation norm. Given two Borel spaces X and Y, $P(Y \mid X)$ stands for the set of all stochastic kernels $\psi(dy \mid x)$ on Y given X; that is, $\psi(dy \mid x) \in P(Y \mid X)$ if $\psi(\cdot \mid x)$ is a probability measure on Y for each $x \in X$, and $\psi(B \mid \cdot)$ is a measurable function on X for each $B \in \mathscr{B}(Y)$.

The Decision Model. We consider the standard (stationary) Markov decision model (X, A, q, c), with state space X, action set A, transition law q, and one-stage cost function c. Both X and A are assumed to be Borel spaces. To each state $x \in X$ we associate a nonempty measurable subset $A(x)$ of A, whose elements are the admissible actions when the system is in state x, and we assume that the set $K := \{(x, a) \mid x \in X, a \in A(x)\}$ of feasible state-action pairs is a measurable subset of $X \times A$. We will also assume that $A(x)$, $c(x, a)$, and $q(dy \mid x, a)$ satisfy the following:

Assumption 2.1. (a) $A(x)$ is a compact set for every $x \in X$.

(b) $c(x, a) \in B(K)$ and, for each $x \in X$, $c(x, a)$ is a lower semi-continuous (l.s.c.) function in $a \in A(x)$.

(c) The transition law $q \in P(X \mid K)$ is such that $\int_X v(y)\, q(dy \mid x, a)$ is l.s.c. in $a \in A(x)$ for each $x \in X$ and $v \in B(X)$.

Control Policies. A *policy* is a sequence $\delta = \{\delta_t\}$ such that, for each $t = 0, 1, \ldots, \delta_t$ is a stochastic kernel on A given the set H_t of histories $h_t := (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$ with $(a_n, x_n) \in K\ \forall n$. Here, x_n and a_n denote the state and action at time n, respectively, and it is assumed that δ_t satisfies the constraint $\delta_t(A(x_t) \mid h_t) = 1$. The class of all policies is denoted by Δ.

Let Φ be the set of all stochastic kernels $\phi \in P(A \mid X)$ such that $\phi(A(x) \mid x) = 1$ for all $x \in X$, and let F be the set of all measurable functions $f\colon X \to A$ such that $f(x) \in A(x)$ for all $x \in X$.

A policy $\delta = \{\delta_t\}$ is said to be a *randomized stationary policy* if there exists $\phi \in \Phi$ such that $\delta_t(\cdot \mid h_t) = \phi(\cdot \mid x_t)$ for every history $h_t = (x_0, a_0, \ldots, x_t) \in H_t$ and $t = 0, 1, \ldots$. In this case we identify δ with $\phi \in \Phi$; in other words, we identify Φ with the set of randomized stationary policies.

Finally, a randomized stationary policy $\phi \in \Phi$ is called (pure or deterministic) *stationary* if there exists $f \in F$ such that $\phi(\{f(x)\} \mid x) = 1$ for all $x \in X$. In such a case, we identify ϕ with $f \in F$, so that F becomes the set of (pure or deterministic) stationary policies.

Notation. Given a randomized stationary policy $\phi \in \Phi$, we write, for $x \in X$,

$$c(x, \phi) := \int_A c(x, a)\, \phi(da \mid x) \qquad \text{and} \qquad q(\cdot \mid x, \phi) = \int_A q(\cdot \mid x, a)\, \phi(da \mid x). \tag{1}$$

For a stationary policy $f \in F$, these expressions reduce to

$$c(x, f) = c(x, f(x)) \qquad \text{and} \qquad q(\cdot \mid x, f) = q(\cdot \mid f(x)),$$

respectively. As is well known, when using a policy $\phi \in \Phi$, the state process $\{x_t\}$ is a Markov chain with stationary transition kernel $q(\cdot \mid x, \phi)$.

Performance Criteria. Let P_x^δ be the induced probability measure when using the policy $\delta \in \Delta$ given the initial state $x_0 = x$ (see, e.g., Hinderer [8, p. 80], for a construction of P_x^δ); the corresponding expectation operator is denoted by E_x^δ.

For any positive integer n, $\delta \in \Delta$ and $x \in X$, let

$$V_n(\delta, x) := \sum_{t=0}^{n-1} E_x^\delta c(x_t, a_t) \qquad n = 1, 2, \ldots, \quad (V_0(\cdot, \cdot) := 0)$$

be the *expected total n-stage cost* under δ when the initial state is x. The corresponding *optimal n-stage cost* is $v_n(x) := \inf_\delta V_n(\delta, x)$. Following Flynn [2, 3], we define the *opportunity cost* of δ at x as

$$O(\delta, x) := \limsup_n \left[V_n(\delta, x) - v_n(x) \right], \tag{2}$$

and δ is said to have *finite opportunity cost* if $O(\delta, \cdot)$ is finite-valued. We also define the usual long-run expected *average cost* per unit time as

$$J(\delta, x) := \limsup_n \left[n^{-1} V_n(\delta, x) \right]$$

and the *optimal average cost* $J(x) := \inf_\delta J(\delta, x)$, $x \in X$.

DEFINITION 2.2. A policy δ^* is said to be

- *average optimal* (AO) if $J(\delta^*, x) = J(x)$ $\forall x \in X$;
- *strong average optimal* (strong AO) if $\limsup_n n^{-1}[V_n(\delta^*, x) - v_n(x)] = 0$.

In this paper, we are specifically interested in the concept of average optimality in the sense of Definition 2.2 and, as already noted by Flynn [2, 3], it is clear that a policy δ is AO if it is strong AO, and the latter in turn is implied if δ has finite opportunity cost. The converse implications, however, do not hold in general, and one of our objectives is to see how strong optimality and finiteness of the opportunity cost relate to the conditions to be stated in Section 3.

3. ERGODICITY AND OPTIMALITY CONDITIONS

In this section we introduce some ergodicity and optimality conditions, and in Section 4 we study some relations between them. A subscript d (d for deterministic) will be used to indicate that a given condition is restricted to the set of (pure or deterministic) stationary policies F.

Ergodicity Conditions

C1. There exists a scalar $\alpha \in (0, 1)$ such that $\|q(\cdot \mid x, \phi) - q(\cdot \mid x', \phi')\| \leqslant 2\alpha$ for all $x, x' \in X$ and $\phi, \phi' \in \Phi$.

C2 (Geometric ergodicity). There exist scalars $\alpha \in (0, 1)$ and $b > 0$ for which the following holds: For each $\phi \in \Phi$ there is a probability measure p_ϕ on X such that

$$\|q^t(\cdot \mid x, \phi) - p_\phi(\cdot)\| \leqslant b\alpha^t \qquad \forall x \in X, \quad \text{and} \quad t = 0, 1, ...,$$

where $q^t(B|x, \phi) = P_x^{\phi}(x_t \in B)$, $B \in \mathscr{B}(X)$, denotes the t-step transition measure when using the policy $\phi \in \Phi$; cf. (1).

C3 (Positive recurrence). For each $\phi \in \Phi$, there exists an invariant probability measure p_ϕ for $q(\cdot \mid \cdot, \phi)$; that is, $p_\phi(B) = \int_X q(B \mid x, \phi)\, p_\phi(dx)$ for all $B \in \mathscr{B}(X)$.

Remark 3.1. C1 implies C2 (with $b = 2$), C3, and also the optimality condition C4 below; see, e.g., [4; 6; 5, p. 57]. Some sufficient conditions for C1 are given in the latter references; they are easily verified in some inventory/production systems as well as in some control of water reservoir problems [11, 12].

Remark 3.2. C1 can be written in several equivalent forms when the state space X is a countable set or $X = R^n$. For instance, suppose that $X = R^n$ and the transition law $q(B \mid x, a)$ has a density $p(y \mid x, a)$ with respect to Lebesgue measure $m(\cdot)$; that is, $q(B \mid x, a) = \int_B p(y \mid x, a)\, dy$ for all $B \in \mathscr{B}(X)$ and $(x, a) \in K$. Then, by Scheffe's Theorem (see, e.g., [5, p. 125]) and using that $|s - t| = s + t - 2\min[s, t]$, we can write

$$\|q(\cdot \mid x, a) - q(\cdot \mid x', a')\| = \int |p(y \mid x, a) - p(y \mid x', a')|\, dy$$

$$= 2 - 2 \int \min[p(y \mid x, a), p(y \mid x', a')]\, dy. \quad (3)$$

(This relation also holds when X is a countable set: replace integrals by sums.) As an example, we can show that Yamada's [12] condition (A1) implies C1. Indeed, consider [12]:

(A1) $X = R^n$, $A = R^m$, and there exists a scalar $\varepsilon > 0$ and a Borel set $C \in \mathscr{B}(X)$ such that $p(y \mid x, a) \geqslant \varepsilon$ for all $y \in C$, $(x, a) \in K$, and $0 < \varepsilon \cdot m(C) < 1$.

Under (A1), $q(\cdot \mid x, \phi)$, with $\phi \in \Phi$, has a density $p(y \mid x, \phi) = \int_A p(y \mid x, a)\, \phi(da \mid x)$ (cf. (1)) satisfying $p(y \mid x, \phi) \geqslant \varepsilon$ for all $y \in C$ and $x \in X$, and (3) yields, for any ϕ and $\phi' \in \Phi$,

$$\|q(\cdot \mid x, \phi) - q(\cdot \mid x', \phi')\| = 2 - 2 \int \min[p(y \mid x, \phi), p(y \mid x', \phi')]\, dy$$

$$\leqslant 2 - 2 \int_C \min[p(y \mid x, \phi), p(y \mid x', \phi')]\, dy$$

$$\leqslant 2(1 - \varepsilon \cdot m(C)),$$

so that C1 holds with $\alpha = 1 - \varepsilon \cdot m(C)$.

Remark 3.3. For the results in Section 4, the geometric ergodicity condition C2 can be replaced by the following: For each $\phi \in \Phi$, there exists a probability measure p_ϕ on X such that

$$\|q^t(\cdot \mid x, \phi) - p_\phi(\cdot)\| \leqslant \beta_t \qquad \text{for all} \quad x \in X,\ t = 0, 1, ..., \tag{4}$$

where $\{\beta_t\}$ is a sequence of constants independent of x and ϕ, and such that $\sum_t \beta_t < \infty$. Sufficient conditions for (4), as well as for C2 and C3, are given, e.g., in [7, 10].

Optimality Conditions

C4. There is a constant j^* and a function $v^* \in B(X)$ such that $(j^*, v^*(\cdot))$ is a solution to the *Optimality Equation*

$$j^* + v^*(x) = \min_{a \in A(x)} \left\{ c(x, a) + \int v^*(y)\, q(dy \mid x, a) \right\}, \qquad x \in X. \tag{5}$$

Equivalently, there is a constant j^* and a function $v^* \in B(X)$ such that $(j^*, v^*(\cdot))$ is an optimal solution to the *problem* (P):

Maximize λ s.t.

$$\lambda + v(x) - \int v(y)\, q(dy \mid x, a) \leqslant c(x, a) \qquad \forall (x, a) \in K, \tag{6}$$

where $\lambda \in R$ and $v \in B(X)$.

C5. There exists $\phi^* \in \Phi$ and $p^* \in P(X)$ such that (ϕ^*, p^*) is an optimal solution to the *dual problem* (D):

Minimize $\int_X \int_A c(x, a)\, \phi(da \mid x)\, p(dx)$ s.t.

$$\int_X \int_A q(B \mid x, a)\, \phi(da \mid x)\, p(dx) = p(B) \qquad \forall B \in \mathscr{B}(X), \tag{7}$$

where $\phi \in \Phi$, $p \in P(X)$.

If we restrict problem (D) to (deterministic) stationary policies $f \in F$, the corresponding "deterministic" version of problem (D) is *problem* (D_d):

Minimize $\int_X c(x, f)\, p(dx)$ s.t.

$$\int_X q(B \mid x, f)\, p(dx) = p(B) \qquad \forall B \in \mathscr{B}(X) \tag{8}$$

where $f \in F$, $p \in P(X)$.

C6. There is a policy $\delta \in \Delta$ with finite opportunity cost.

Notice that problem (P) is "linear" in $(\lambda, v(\cdot))$, whereas (D) (or (D_d)) is nonlinear in (ϕ, p). However, if the transition law q is absolutely continuous with respect to some sigma-finite measure μ on X—e.g., $\mu = m =$ Lebesgue measure if $X = R^n$ (cf. Remark 3.2 or [12]), or $\mu =$ counting measure if X is a denumerable set (cf. [1])—then (D) can be written as the standard dual linear problem for (P), as in Linear Programming.

Remark 3.4. We can also write the optimality equation (5) as $\min_{a \in A(x)} D(x, a) = 0$, where

$$D(x, a) := c(x, a) + \int v^*(y)\, q(dy \mid x, a) - j^* - v^*(x), \qquad (x, a) \in K,$$

is the so-called "discrepancy" function. Let $F^* := \{f \in F \mid D(x, f(x)) = 0\}$; that is, $f \in F^*$ if $f(x) \in A(x)$ minimizes the right hand side (r.h.s.) of (5) for all $x \in X$. Under Assumption 2.1, well-known Measurable Selection theorems imply that F^* is nonempty. On the other hand, if C4 holds, then j^* is the optimal cost function, i.e., $j^* = J(x)$ for all $x \in X$, and moreover, $j^* = J(f, x)$ if $f \in F^*$, so that $f \in F^*$ is AO. We will show in Theorem 4.2 that a stationary policy $f \in F^*$ is in fact *strong* AO (Definition 2.2).

4. Theorems

The objective in this section is to prove some results connecting conditions C4, C5, and C6. Theorem 4.1 is a "duality theorem": it gives conditions under which the existence of an optimal solution to the "primal" problem (P) in C4 yields an optimal solution to the "dual" problem (D)—or to the deterministic version (D_d)—in C5, and conversely. Theorem 4.2 shows that C4 implies C7, which extends to our present Borel-space setting a result of Flynn [2] when X is a denumerable set and A is finite.

THEOREM 4.1. (a) *Suppose the ergodicity condition* C3 *holds. Then*:

(i) *the problems* (P), (D) [*and* (D_d)] *in* C4 *and* C5, *respectively, are feasible*;

(ii) *for any feasible solutions* $(\lambda, v(\cdot))$ *of* (P) *and* (ϕ, p) [*respectively* (f, p)] *of* (D) [*respectively* (D_d)],

$$\lambda \leqslant \int_X \int_A c(x, a)\, \phi(da \mid x)\, p(dx) \qquad \left[\textit{respectively } \lambda \leqslant \int_X c(x, f)\, p(dx)\right]; \quad (9)$$

(iii) *if* (P) *has an optimal* (*bounded*) *solution, then so do* (D) *and* (D_d), *and the optimal values of the corresponding objective functions are* *equal*; *in fact, an optimal solution to* (D) *can be chosen from the set of optimal solutions to* (D_d). (*See also Remark* 4.3.)

(b) *If* C2 *holds and* (D) [*or* (D_d)] *has an optimal solution, then so does* (P) *and the corresponding optimal values of* (P) *and* (D) [*or* (D_d)] *are equal.*

Proof. (a) (cf. [12]) (i) To see that (P) is feasible it suffices to take $v(\cdot)=0$ and λ sufficiently small. Feasibility of (D) or (D_d) follows from C3: if $\phi \in \Phi$ and p_ϕ is an invariant probability measure for $q(\cdot \mid \cdot, \phi)$, then the pair (ϕ, p) satisfies (7). Similarly, if $f \in F$, then (f, p_f) satisfies (8).

(ii) Now suppose that $(\lambda, v(\cdot))$ satisfies (6) and (ϕ, p) satisfies (7). Then, integrate (6) with respect to $\phi(da \mid x)$ and then with respect to $p(dx)$ to obtain

$$\lambda + \int_X v\,dp - \int_X \int_A \left[\int_X v(y)\,q(dy \mid x, a)\right] \phi(da \mid x)\,p(dx)$$
$$\leqslant \int_X \int_A c(x, a)\,\phi(da \mid x)\,p(dx).$$

Finally, using Fubini's theorem and (7), the third term reduces to $\int_X v\,dp$ so that the latter inequality reduces to (9). The proof for (f, p), satisfying (8) is similar.

(iii) Let $(j^*, v^*(\cdot))$ be a (bounded) solution to (5) and take $f \in F^*$; that is,

$$j^* + v^*(x) = c(x, f) + \int v^*(y)\,q(dy \mid x, f), \qquad x \in X. \tag{10}$$

Now let $\phi^* \in \Phi$ be such that $\phi^*(\cdot \mid x)$ is the probability measure concentrated at $f(x)$ $\forall x \in X$, and let $p^* = p_f$ be a corresponding invariant probability measure. Then (10) can be written as

$$j^* + v^*_t(x) = c(x, \phi^*) + \int v^*(y)\,q(dy \mid x, \phi^*),$$

and integration with respect to $p^*(dx)$ yields $j^* = \iint c(x, a)\ \phi^*(da \mid x)\ p^*(dx)$, which yields the desired conclusion (cf. Remark 3.4).

(b) Let (ϕ^*, p^*) be an optimal solution to (D), and define

$$\begin{aligned} j^* &:= \int_X \int_A c(x, a)\,\phi^*(da \mid x)\,p^*(dx) = \int_X c(x, \phi^*)\,p^*(dx) \\ v^*(x) &:= \sum_{t=0}^{\infty} E_x^{\phi^*}[c(x_t, \phi^*) - j^*]. \end{aligned} \tag{11}$$

Under condition C2,

$$|E_x^{\phi^*} c(x_t, \phi^*) - j^*| = \left| \int c(y, \phi^*)[q^t(dy \mid x, \phi^*) - p^*(dy)] \right|$$
$$\leqslant \|c\| \cdot \|q^t(\cdot \mid x, \phi^*) - p^*(\cdot)\| \leqslant b\|c\| \alpha^t,$$

and therefore, v^* is uniformly bounded in $x \in X$: $|v^*(x)| \leqslant b\|c\|/(1-\alpha)$. On the other hand, by definition of v^* and the Markov property,

$$v^*(x) = c(x, \phi^*) - j^* + \sum_{t=1}^{\infty} E_x^{\phi^*}[c(x_t, \phi^*) - j^*]$$
$$= c(x, \phi^*) - j^* + \int v^*(y)\, q(dy \mid x, \phi^*);$$

that is,

$$j^* + v^*(x) = c(x, \phi^*) + \int v^*(y)\, q(dy \mid x, \phi^*) \qquad \forall x \in X, \tag{12}$$

so that $(j^*, v^*(\cdot))$ is feasible for (P). To show that $(j^*, v^*(\cdot))$ is optimal for (P), first note that, from (12),

$$j^* + v^*(x) \geqslant \min_{a \in A(x)} \left\{ c(x, a) + \int v^*(y)\, q(dy \mid x, a) \right\} \qquad \forall x \in X. \tag{13}$$

Now let $\bar{f} \in F$ be a minimizer of the r.h.s. of (13), so that

$$j^* + v^*(x) \geqslant c(x, \bar{f}) + \int v^*(y)\, q(dy \mid x, \bar{f}), \qquad x \in X.$$

Iteration of this inequality yields

$$nj^* + v^*(x) \geqslant \sum_{t=0}^{n-1} E_x^{\bar{f}}[c(x_t, \bar{f})] + E_x^{\bar{f}} v^*(x_n),$$

so that, dividing over n, taking the limit as $n \to \infty$, and using the boundedness of v^*, $j^* \geqslant J(\bar{f})$, where $J(\bar{f}) = \int c(y, \bar{f})\, p_{\bar{f}}(dy) = J(\bar{f}, x)\ \forall x \in X$, by C2. On the other hand, the optimality of (ϕ^*, p^*) and the definition (11) of j^* imply that $j^* \leqslant J(\bar{f})$. Hence $j^* = J(\bar{f})$ and the equality holds in (13), i.e., (j^*, v^*) satisfies the optimality equation (5). Clearly, the above arguments still work if, instead of an optimal solution to (D), we take an optimal solution (f^*, p^*) to (D_d). ∎

In the proof of Theorem 4.2 we will use that the optimal n-stage cost functions v_n, $n = 1, 2, ...$, can be written iteratively as

$$v_n(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \int v_{n-1}(y)\, q(dy \mid x, a) \right\}, \qquad x \in X, \tag{14}$$

with $v_0 := 0$; see, e.g., [5, 8]. Also recall the definitions of $D(x, a)$ and F^* in Remark 3.4.

THEOREM 4.2. C4 *implies* C6; *more precisely, if* C4 *holds and* $f^* \in F^*$, *then* f^* *has finite opportunity cost, and therefore,* f^* *is strong AO.*

Proof. Let $(j^*, v^*(\cdot))$ be a bounded solution to (5) and let $f^* \in F^*$. We wish to estimate the opportunity cost $O(f^*, x)$ in (2).

Let us define $e_n(x) := v_n(x) - v^*(x) - nj^*$, for $x \in X$, $n = 0, 1, \ldots$. Notice that $e_0(x) = -v^*(x)$. We will first show that $\|e_n\|$ is non-increasing, i.e.,

$$\|e_{n+1}\| \leqslant \|e_n\|, \qquad \forall n = 0, 1, \ldots, \tag{15}$$

so that $\|e_n\| \leqslant \|e_0\| = \|v^*\| < \infty$ for all n. To begin, a direct calculation using (14) yields

$$e_{n+1}(x) = \min_{a \in A(x)} \left\{ D(x, a) + \int e_n(y)\, q(dy \mid x, a) \right\}, \qquad x \in X, \tag{16}$$

where D is the "discrepancy" function in Remark 3.4. Thus, if we take $\bar{f} \in F^*$, then $D(x, \bar{f}(x)) = 0$ for all $x \in X$, and from (16),

$$e_{n+1}(x) \leqslant \int e_n(y)\, q(dy \mid x, \bar{f}) \leqslant \|e_n\| \qquad \forall x \in X.$$

On the other hand, $D(x, a) \geqslant 0$ implies

$$e_{n+1}(x) \geqslant \min_{a \in A(x)} \int e_n(y)\, q(dy \mid x, a) \geqslant -\|e_n\|, \qquad \forall x \in X.$$

Thus $|e_{n+1}(x)| \leqslant \|e_n\|$ $\forall x \in X$, and (15) follows.

Now if $f^* \in F^*$, (5) becomes

$$j^* + v^*(x) = c(x, f^*) + \int v^*(y)\, q(dy \mid x, f^*) \qquad \forall x \in X,$$

and iterating, we obtain

$$V_n(f^*, x) := \sum_{t=0}^{n-1} E_x^{f^*} c(x_t, f^*) = v^*(x) + nj^* - E_x^{f^*} v^*(x_n).$$

Therefore,

$$0 \leqslant V_n(f^*, x) - v_n(x) = -e_n(x) - E_x^{f^*} v^*(x_n) \leqslant 2\|v^*\| \qquad \forall x \in X,$$

and by (2), $O(f^*, x) \leq 2\|v^*\|$ $\forall x \in X$. This completes the proof of Theorem 4.1. ∎

Combining Remark 3.1 and Theorems 4.1 and 4.2, we obtain other sufficient conditions for C6:

COROLLARY. (a) C1 *implies* C6.

(b) C2 *and* C5 [*or* C5$_d$, *i.e., replacing* (D) *by* (D$_d$)] *together imply* C6.

More generally, C6 is implied by any set of sufficient conditions for C4, which in turn can be obtained in a number of ways [2, 4, 5, 7, 9].

Remark 4.3. As can be seen in the proof of Theorem 4.1, the conclusion in part (a)(iii) of that theorem still holds if C3 is replaced by the following weaker condition: [(P) has an optimal bounded solution and] *there exists a stationary policy* $f \in F^*$ *such that* $q(\cdot \mid \cdot, f)$ *has an invariant probability measure.*

REFERENCES

1. J. A. FILAR AND T. A. SCHULTZ, Communicating MDPs: equivalence and LP properties, *Oper. Res. Lett.* **7** (1988), 303–307.
2. J. FLYNN, On optimality criteria for dynamic programs with long finite horizons, *J. Math. Anal. Appl.* **76** (1980), 202–208.
3. J. FLYNN, Optimal steady states, excessive functions, and deterministic dynamic programs, *J. Math. Anal. Appl.* **144** (1989), 586–594.
4. J. P. GEORGIN, Contrôle de chaînes de Markov sur des espaces arbitraires, *Ann. Inst. H. Poincaré* **14** (1978), 255–277.
5. O. HERNÁNDEZ-LERMA, "Adaptive Markov Control Processes," Springer-Verlag, New York, 1989.
6. O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, A forecast horizon and a stopping rule for general Markov decision processes, *J. Math. Anal. Appl.* **132** (1988), 388–400.
7. O. HERNÁNDEZ-LERMA, R. MONTES-DE-OCA, AND R. CAVAZOS-CADENA, Recurrence conditions for Markov decision processes with Borel state space, *Ann. Oper. Res.* (1991), to appear.
8. K. HINDERER, "Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter," Lecture Notes Oper. Res., Vol. 33, Springer-Verlag, New York, 1970.
9. M. KURANO, The existence of a minimum pair of state and policy for Markov decision processes under the hypothesis of Doeblin, *SIAM J. Control Optim.* **27** (1989), 296–307.
10. R. L. TWEEDIE, Criteria for rates of convergence of Markov chains, with applications to queueing and storage theory, *in* "Papers in Probability Statistics and Analysis" (J. F. C. Kingman and G. E. H. Reuter, Eds.), pp. 260–276, Cambridge Univ. Press, Cambridge, 1983.
11. S. YAKOWITZ, Dynamic programming applications in water resources, *Water Resources Res.* **18** (1982), 673–696.
12. K. YAMADA, Duality theorem in Markovian decision problems, *J. Math. Anal. Appl.* **50** (1975), 579–595.